

Posudek dizertační práce

Název dizertační práce

Rozsáhlé informační systémy – standardizace metadat

Instituce

Západočeská univerzita v Plzni, Fakulta aplikovaných věd

Autor

Ing. Petr Včelák

Školitel

doc. Dr. Ing. Jana Klečková

Oponent

doc. RNDr. Petr Šaloun, Ph.D.

Přehled a zhodnocení významu disertační práce pro obor

Posuzovaná dizertační práce, jejímž autorem je Ing. Petr Včelák, má celkem 180 stran textu včetně příloh a je napsána česky. Značnou část práce představují přílohy, seznam zkratek a často používaných pojmů, standardní přehledy obrázků, tabulek, výpisů a také použitá literatura. Osobně považuji 97 stran textu práce za zvládnutí formátu.

Dizertační práce představuje původní výzkum doktoranda v uznávané a stále se rozvíjející se výzkumné oblasti. Potřebnost metadat v medicínských rozsáhlých informačních systémech a datových výzkumech je evidentní a jejich správný nejen návrh, ale i implementace a provozování má praktický význam.

Vyjádření k postupu řešení problému, použitým metodám a splnění cíle

Dizertační práce je sepsána správně jako kniha, dovolující sekvenční čtení od začátku do konce. Uvádí definice základních pojmů, jakými jsou například definice ontologie, která je citována správně z původního zdroje, jímž je populárně-vědecký Scientific American, viz [180] v práci. Na straně 16 je rovněž správně citována práce Grubera, v práci [83], ale není tam uvedena ona pověstná zásadní definice: „An ontology is a formal explicit specification of a shared conceptualization”. Byť je ontologie na uvedeném místě popsána dostatečně přesně.

Cíle dizertace jsou:

- Navrhnout standardizované procesy pro podporu medicínského výzkumu založeného na primárních zdravotnických datech.
- Doménová ontologie DASTA Ontology, určená k sekundárnímu využití zdravotnických dat používaných v České republice.
- Návrh a implementace doménových ontologií pro metadata o anamnéze, diagnostice, terapii nebo dispenzarizaci u pacientů po cévní mozkové příhodě...

Navíc byla vytvořena Population Ontology pro možnost určení incidencí.

Vyjádření formální úpravě a jazykové úrovni disertační práce

Práce je sepsána správnou odbornou češtinou, je vytvořena systematicky a postupně buduje pojmový, znalostní i datový aparát a procesy potřebné ke komplexnímu mezioborovému řešení stanovených cílů. Po formální stránce hodnotím práci jako kvalitní.

Výsledky disertační práce a k původním přínos předkladatele

Součástí datově orientovaného výzkumu doktoranda bylo vytvoření komplexní platformy Medical Research & Education (MRE). Tento rozsáhlý informační systém je v práci odpovídajícím

způsobem navržen, popsán na úrovni standardizace procesů, schématu metadat, tvorby ontologií a slovníků, technologií i specifik, které sebou nese rozsáhlá mezioborovost navrženého řešení. Doktorand se rozsáhle věnuje ontologiím, navrženým pro samostatné „případové studie“ pro použití například v rámci popisu cévních mozkových příhod, ochraně osobních údajů a dalších, jak je popsáno v práci. Výzkumná platforma byla nejen navržena, ale je i implementována a provozována a metadata jsou spravována v rozsáhlém online úložišti. Dosažené výsledky byly mj. součástí mezinárodní grantové přihlášky v rámci výzvy H2020 a jsou rozsáhle diskutovány v závěru práce, kde jsou následně shrnuty přínosy práce a doporučení.

Vyjádření k publikacím studenta

Sumarizace výstupů doktoranda ve špičkových infromaticko-lékařských časopisech hodnocených v Q1, které jsou již i odpovídajícím způsobem citovány je vynikajícím dokladem jeho výzkumného přínosu i dlouhodobého zapojení do týmové práce. Podobné je to s konferenčními příspěvky v kvalitativní úrovni WoS a Scopus. Zpravidla doktorand publikoval společně se svou školitelkou, která se tak spolupodílela na jeho odborném růstu a projektovém zapojení a disseminaci jejich výsledků. Jak jsem již zmínil, publikace doktoranda jsou odpovídajícím způsobem citovány. Naprosto není pochyb o kvalitě jeho výsledků a jejich přijetí odbornou veřejností.

Diskuse, otázky, náměty a poznámky pro autora

Shodou okolností se podílím na mezioborovém projektu TAČR „Výzkum a vývoj podpůrných sítí a informačních systémů pro neformální pečující o osoby po cévních mozkových příhodách“. Jaká doporučení byste mohl dát odborné komunitě spolupracujících sociálních, rehabilitačních, lékařských pracovníků a IT specialistů, aby při vzájemné komunikaci i podpoře laických uživatelů mohli postupně přejít na využití metadat a dalších výsledků Vaší práce?

Shrnutí

Disertační práce obsahuje přehled stavu poznání v dané výzkumné oblasti, přehled dosavadních publikovaných výsledků založený zejména na výzkumu rozsáhlých informačních systémů v mezioborové oblasti lékařského výzkumu a standardizaci metadat v ní. Práce obsahuje původní výsledky výzkumu a ucelenou řadu ontologií, metodik a doporučení. Výsledky doktoranda byly publikovány na mezinárodní i národní úrovni na odpovídajících odborných a vědeckých časopisech a konferencích a lze tedy konstatovat, že výsledky jsou kvalitní, a že s jádrem původního výzkumu doktoranda byla odborná veřejnost dostatečně seznámena.

Ing. Petr Včelák v disertační práci prokázal schopnost výzkumné individuální i týmové práce. S ohledem na výše uvedené *doporučuji disertační práci k obhajobě před příslušnou komisí a doporučuji udělení akademického titulu „philosophiae doctor“* jejímu autorovi.

V Náměšti na Hané dne 29. 5. 2021

doc. RNDr. Petr Šaloun, Ph.D.
katedra informatiky
FEI VŠB-Technická univerzita Ostrava
17. listopadu 15
708 33 Ostrava-Poruba
e-mail: petr.saloun@vsb.cz

Oponentský posudek disertační práce

Studijní program: *Informatika a výpočetní technika*

Jméno studenta: *Ing. Petr Včelák*

Název práce: *Rozsáhlé informační systémy – standardizace metadat*

Význam disertační práce pro obor

Předložená disertační práce se zabývá problematikou související s rozsáhlými informačními systémy pracujícími s velkým množstvím heterogenních dat z pohledu standardizace metadat. Jde o problematiku aktuální, protože takových systémů existuje řada, stejně jako požadavků na jejich interoperabilitu a pro ni potřebnou standardizaci. Autor disertace předkládá řešení pro oblast multidisciplinárního medicínského výzkumu, na kterém participují zdravotnická zařízení poskytující medicínská data a provádějící výzkum společně s výzkumníky z jiných organizací. Standardizace je založena na využití ontologického popisu. Konkrétní aplikaci navrženého přístupu ukazuje na příkladu výzkumné platformy Medical Research and Education (MRE) vyvíjené na pracovišti autora.

Přestože ontologie navrhl a implementoval pro medicínská data, případně i jejich užší specializace (cévní mozkové příhody, idiopatické střevní záněty), a procesy standardizoval pro potřeby platformy MRE, lze použitý přístup i některé dílčí implementace uplatnit i v jiných oblastech, kde je požadována interoperabilita systémů pracujících s daty podobného charakteru.

Postup řešení problému, použité metody, splnění cíle

Při zpracování autor použil základní postup běžný pro disertační práce – vytyčení cílů disertace, analýzu současného stavu v oblasti zaměření disertace, návrh řešení, ověření a vyhodnocení řešení.

Práce je členěna do 12 kapitol. První kapitola je úvodem do problematiky a deklaruje také tři cíle disertace. Prvním je návrh standardizovaných procesů pro podporu medicínského výzkumu založeného na primárních zdravotnických datech. Druhým je návrh a implementace doménové ontologie pro sekundární využití zdravotnických dat v České republice a třetí návrh a implementace ontologií pro rozšířená metadata týkající se pacientů po cévní mozkové příhodě a pacientů s idiopatickými střevními záněty.

Druhá až pátá kapitola jsou věnovány podrobnějšímu uvedení do problematiky a shrnutí současného stavu řešení problematiky jak ve světě, tak v České republice, pro kterou existují určitá specifika.

Druhá kapitola shrnuje některé důležité pojmy související s informačními systémy používané v dalších kapitolách. Podobně třetí kapitola shrnuje důležité pojmy související s metadaty, vysvětluje jejich roli a některé další vlastnosti, včetně způsobů jejich uložení. Důležitou součástí třetí kapitoly je část týkající se ontologií a poměrně podrobné zhodnocení stavu standardizace metadat. Je zřejmé, že autor má v této oblasti velmi dobrý přehled.

Čtvrtá kapitola se zaměřuje podrobněji na technologie sémantického webu, na kterých je řešení standardizace metadat prezentované v disertaci založeno. I obsah této části svědčí o přehledu, včetně přehledu o dostupné literatuře a jiných zdrojích, kde jsou podrobnější informace.

Pátá kapitola je věnována výzkumu v medicíně z pohledu zdravotnických dat a standardizace jejich formátu pro účely výměny informací a jejich strojového zpracování. Opět s odkazy na řadu použitých informačních zdrojů prezentuje a hodnotí autor stav u nás v kontextu trendů v Evropské unii, případně jinde ve světě. Zaměřuje se rovněž na standardy pro obrazová a signálová data. Důležitým aspektem zdravotnických dat z hlediska jejich použití je ochrana osobních údajů. Proto se v této části věnuje i požadavkům v této oblasti. Dále zde stručně prezentuje dvě specifická rozšíření obecných medicínských dat, na která se ve své práci také zaměřuje, a to data pacientů po cévní mozkové příhodě a pacientů s idiopatickými střevními záněty. V závěru páté kapitoly ještě stručně představuje některé jak doménově specializované, tak obecně použitelné registry a systémy využitelné pro správu medicínských dat.

Šestá až desátá kapitola jsou věnovány vlastním výsledkům autora. V šesté kapitole nazvané „Použité metody“ autor uvádí jednak některé obecně použitelné metody a přístupy, které při řešení využil, jednak fáze životního cyklu ontologií a pravidla tvorby ontologií a slovníků.

Obsahem sedmé kapitoly je popis navržených ontologií, naplňuje tedy druhý a třetí cíl disertace. Popis každé z navržených ontologií má stejnou strukturu, která sestává ze tří částí vycházejících z fázi životního cyklu ontologie – studie proveditelnosti a analýza domény; konceptualizace a implementace; diskuse, údržba a užití. Slovní popis konceptualizace je vždy doplněn o zjednodušený diagram ontologie používající zřejmě notace UML. Vždy je uveden i odkaz na kompletní ontologii v OWL na webových stránkách <https://mre.zcu.cz/>.

Osmá kapitola obsahuje výsledky vztahující se k prvnímu cíli disertace – standardizaci procesů pro podporu medicínského výzkumu založeného na primárních zdravotnických datech. Standardizace je provedena v rámci platformy MRE, proto je nejprve vysvětlena celková koncepce řešení platformy a popsány její hlavní komponenty. Následně je prezentován proces pro zpracování nových (primárních) zdravotnických dat, jehož důležitou součástí je zajištění ochrany osobních údajů, a proces pro operace ETL se sekundárními daty při využití ontologií, tzv. ETL proces. U obou procesů je popsán software vytvořený na podporu některých klíčových úloh (zajištění ochrany osobních údajů, extraktory pro software MetaMed). Dále jsou stručně popsány některé další softwarové nástroje v podobě knihoven či aplikací vytvořených na podporu vývoje aplikací využívajících platformu MRE a na podporu práce uživatelů s daty uloženými v archívu platformy MRE (dotazování v jazyce SPARQL, generování webových formulářů).

Devátá kapitola prezentuje výsledky ověření navrženého řešení. Ověření vychází z použití platformy MRE v několika projektech. Jsou zde uvedeny tabulky se statistickými údaji týkajícími se archívu platformy MRE, rychlosti zpracování metadat a importu v RDF a jsou uvedeny některé údaje vztahující se ke kvalitě ochrany osobních údajů. Údaje tabulek i způsob ověření nebo popis zajištění řady dalších vlastností navrženého řešení jsou v textu prezentovány a diskutovány na rozumné úrovni. Metriky ontologií jsou v příloze F a přílohy G, resp. H obsahují seznam klíčových požadavků, resp. hodnocení jejich splnění.

V kapitole deset nazvané „Diskuze“ autor diskutuje výsledky disertace v kontextu provedené rešerše současného stavu, kterou tady ještě doplňuje. Uvádí zde také využití výsledků v některých dalších projektech.

V kapitolách 11 až 13 autor shrnuje ze svého pohledu přínosy práce, prezentuje několik doporučení týkajících se využití výsledků disertací a možného dalšího rozvoje a v závěru konfrontuje výsledky s vytyčenými cíli.

Samotný text práce je doplněn deseti přílohami a DVD. Přílohy doplňují informace v hlavním textu a/nebo ukazují konkrétní příklady použití. Poslední z příloh obsahuje seznam publikací autora, jejich citací a informací o dalších relevantních aktivitách autora. DVD obsahuje text disertace a programovou dokumentaci.

Celkově považují postup řešení i použité metody za vhodné. Deklarovaný cíl disertace považují za splněný – první cíl pokrývá kapitola 8, druhý a třetí kapitola 7. Ověření a diskusi výsledků jsou věnovány kapitoly 9 a 10.

K obsahu mám tři drobné připomínky:

1. Mohlo být uvedeno něco k notaci použité u zjednodušených diagramů ontologií.
2. Detailní informace týkající se struktury adresářů na str. 72 až 74 mohly být v příloze nebo v programové dokumentaci na DVD a v hlavním textu stručně spíše jenom koncepce řešení.
3. Jak vznik a zdůvodnění klíčových požadavků z přílohy G, tak podrobnější informace k hodnocení jejich naplnění v příloze H nejsou v textu ani samotných přílohách blíže vysvětleny.

Výsledky disertační práce, původní přínos předkladatele disertační práce

Za hlavní původní výsledky práce považují navržené ontologie pro standardizaci medicínských metadat a jejich implementaci v rámci platformy MRE a také příspěvek autora ke standardizaci základních procesů souvisejících se zpracováním primárních medicínských dat a s extrakcí, transformací a exportem uložených dat pro sekundární využití. Ve druhém případě jde jak o návrh procesů, tak implementaci podpůrných softwarových nástrojů. Přestože jde o problém specifické domény

medicínského výzkumu, principy a získané poznatky jsou využitelné i v jiných aplikačních oblastech, případně obecně při použití technologií sémantického webu.

Systematika, přehlednost, formální úprava a jazyková úroveň

Práce má logickou strukturu, kapitoly na sebe dobře navazují. Jedinou výhradu mám ke kapitole 8, kde mohlo být v úvodu více informací k základním požadavkům na platformu MRE a kontextu předpokládaného použití (viz obr. 8.4). Odkaz na přílohu G se seznamem požadavků považuji v tomto smyslu pro čtenáře za nedostatečný.

Po formální i jazykové stránce považuji práci za dobře zpracovanou. Kladně hodnotím existenci seznamu zkratk, kterých je v textu poměrně hodně. Jedinou drobnou připomínku mám k nejednotným odkazům na použité zdroje (pořadové číslo ze seznamu literatury vs. řetězec ze seznamu v příloze J). Pro čtenáře disertace v listinné podobě, který nemůže použít aktivní odkaz, je tento způsob bez upozornění poněkud matoucí.

Publikační činnost

Publikace autora jsou uvedeny v příloze J. Je mezi nimi 7 článků v časopisech, 3 z nichž jsou indexované ve WoS a/nebo Scopus. Z pohledu informatického jádra disertace považuji z nich za nejvýznamnější především [VKKK19] (Q1 v kategorii Health Informatics) a [VKR11], kde je Ing. Včelák současně prvním autorem. U ostatních publikací jde zřejmě spíše o využití platformy MRE, případně jiných vytvořených softwarových produktů, při konkrétním medicínském výzkumu.

Dále je autorem nebo spoluautorem 12 příspěvků na konferencích, jejichž sborníky jsou indexovány ve WoS a/nebo Scopus. Z nich prvních šest v příloze J disertace považuji za přímo se vztahující k tématu disertace. Ing. Včelák je u nich rovněž prvním autorem. Několik dalších publikací vztahujících se k tématu disertace je uvedeno v příloze J v části „Ostatní“.

Počet citací jeho publikací je ve WoS 16 (bez autocitací) a ve Scopus 19 (bez autocitací všech autorů).

Celkově hodnotím publikační činnost Ing. Včeláka jako dostatečně kvalitní, odpovídající požadavkům na publikace doktoranda. Jádro disertační práce považuji za dostatečně publikované.

Celkové hodnocení

Ing. Petr Včelák podle mého názoru prokázal svou disertační prací, ostatními svými publikacemi i svým zapojením do výzkumu schopnost vědecké práce v oboru „Informatika a výpočetní technika“. Proto jeho disertační práci k obhajobě doporučuji.

Dotazy k obhajobě:

1. U ontologie MRE Ontology v diskusi na str. 60 uvádíte, že subsystém pro plánování a provedení úloh je nad rámec práce. Znamená to, že podpora pro tuto část ontologie zatím nebyla implementována? Pokud ne, počítá se s tím?
2. Aplikace *PureImage* se produkčně používá pouze pro klasifikaci DICOM snímků. Máte nějaký odhad pokrytí snímků s osobními údaji stávajícím způsobem de-identifikace (pokud existuje), tj. do jaké míry zůstávají osobní údaje neodhaleny?
3. Do jaké míry využíváte nebo se dá využít výsledků iniciativ v oblasti otevřených dat v ČR?

V Brně dne 16. 3. 2021

doc. Ing. Jaroslav Zendulka, CSc.
ÚIFS FIT VUT v Brně

Západočeská univerzita v Plzni

Doručeno: 19.03.2021

ZCU 006332/2021

listy: 3

přílohy:

druh:



zcupes14094c5

